숭실대학교
Soongsil University

# 졸업생이 들려주는
# 데이터 사이언티스트로의 길

2022. 4. 7.

**Presenter: PhD. 고종현**

** 2012.03~2016.02: 숭실대학교 정보통계·보험수리학과 학사*
** 2016.03~2022.02: 울산과학기술원 산업공학과 박사*
** 2022.03~2022.04: 울산과학기술원 4차산업혁신연구소 박사후연구원*
** 2022.05~ : 독일 아헨공대 Process and Data Science 박사후연구원*

# CONTENTS

1. 정보통계 보험수리학과의 고종현 03
- 대학생 시절의 진로 고민: 남들과 다른 길을 간 이유

2. 연구자로서의 길 08
- 대학원을 생각하고 계신 후배님들에게 드리는 조언

3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝 13
- 데이터 사이언스 분야 중, 매우 중요하면서도 미개척된 분야

# 고종현, PhD

2012.03 ~ 2016.02: 숭실대학교 정보통계보험수리학과 (학사)

2016.03 ~ 2022.02: 울산과학기술원 산업공학과 (석박사통합과정)

2021.10 ~ 2021.11: 네덜란드, 위트레흐트 대학 방문연구원

2022.03 ~ 2022.04: 울산과학기술원 4차산업혁신연구소 박사후연구원

2022.05 ~ : 독일 아헨공대 박사후연구원 (프로세스 & 데이터 사이언스 부서)

# MBTI 별 어울리는 직업 (잡코리아 제공)

**ENFP #재기 발랄한 활동가**
: 크리에이티브 디렉터, 디자이너, 시나리오작가,방송 프로듀서, 홍보 컨설턴트, 상담사, 상품 기획자

**ENFJ #정의로운 사회운동가**
: 아나운서, 리포터, 방송 MC, 언어교사, 아동 복지사, CEO, 취업 컨설턴트, 동시 통역가

**INFP #열정적인 중재자**
: 예술가, 소설가, 시인, 음악가, 미술 치료사, 사회복지사, 작곡가, 사서

**INFJ #선의의 옹호자**
: 직업상담사, 특수 교사, 노인 복지사, 아트 디렉터, 프리랜서 기획, 저널리스트, 상품기획 MD

**ENTP #논쟁을 즐기는 변론가**
: 발명가, 벤처 사업가, 에이전트, 배우, 가수, 영화감독, 칼럼니스트, 정치인

**ENTJ #대담한 통솔자**
: 경영 컨설턴트, 공인 중개사, 관리사, 변호사, 재무 상담사, 경제 분석가, 벤처 투자가, 판사

**INTP #논리적인 사색가**
: 경제학자, 심리학자, 경찰, 프로그래머, 천문학자, 비평가, 아트디렉터, 연구원

**INTJ #용의주도한 전략가**
: 분석가, 회계사, 인류학자, 파일럿, 경영 컨설턴트, 제약회사 연구원, 웹 개발자, 최고 재무 책임자

**ESTJ #엄격한 관리자**
: 감독관, 예산 분석가, 은행장, 정책 책임자, 보안 요원, 기관사, 교육 전문가

**ESFJ #사교적인 외교관**
: 홍보 책임자, 호텔 지배인, 마케팅 책임자, 초등학교 교사, 특수 교사, 비서, 유치원 교사

**ISTJ #청렴 결백한 논리주의자**
: 통계학자, 바이어, 기상학자, 법률 연구원, 보험 심사관, 형사, 감정 평가사, 세관 조사관

**ISFJ #용감한 수호자**
: 행정 보조원, 인사 관리자, 신용 상담가, 보호 감찰관, 물리치료사, 정신과 의사, 방사선 기사

**ESTP #모험을 즐기는 사업가**
: 경찰관, 소방관, 군 장교, 펀드 매니저, 은행원, 기자, 여행 가이드, 건축 엔지니어

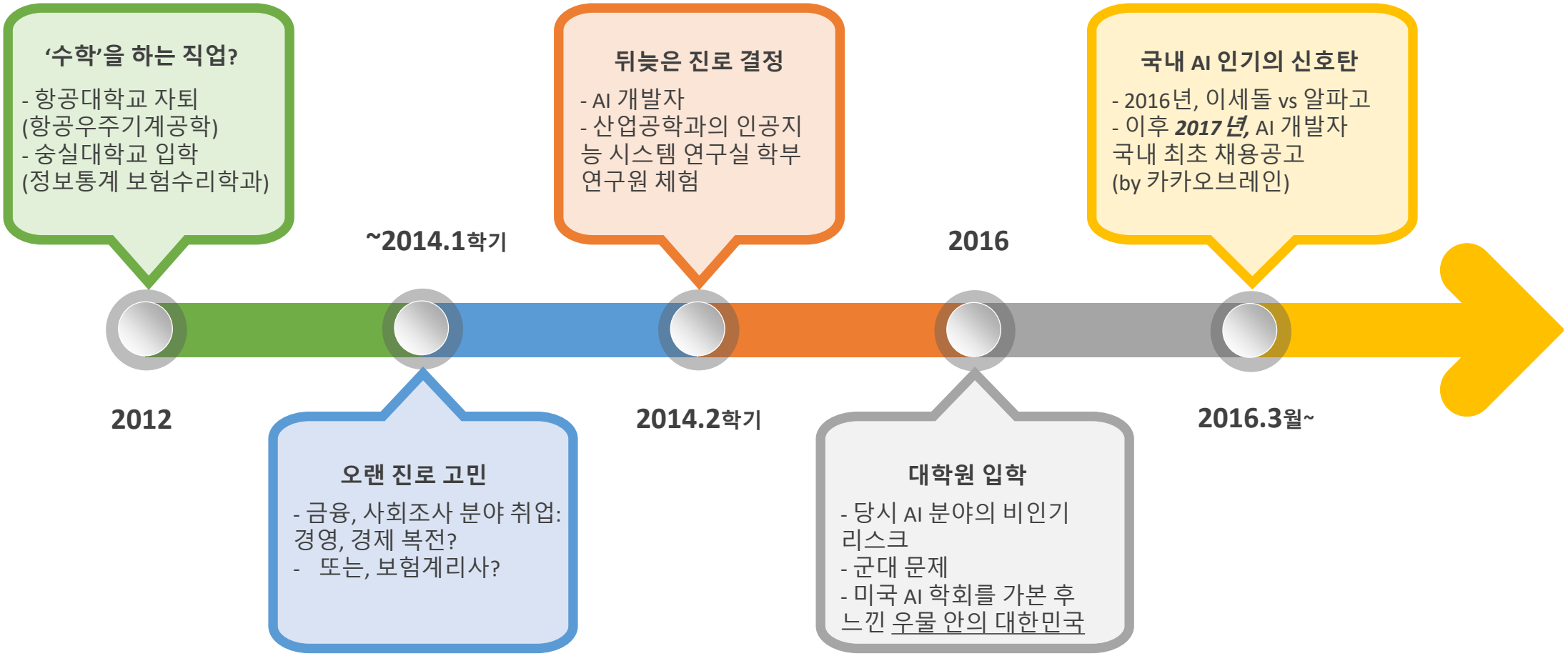
**ESFP #자유로운 영혼의 연예인**
: 코미디언, 의상 디자이너, 일러스트레이터, 애니메이터, 여행 상품 기획자, 놀이 치료사, 승무원

**ISTP #만능 재주꾼**
: 파일럿, 카레이서, 범죄학자, 사진 작가, 판매원, 운동선수, 항공기 정비사, 네트워크 관리자

**ISFP #호기심 많은 예술가**
: 보석 세공사, 음향 디자이너, 만화가, 지질학자, 사육사, 수의사, 법률 비서, 약사

*통계학과 12학번 사례 (3명 모두 ISFJ)
- 통계청 7,9급 공무원: 매우 만족
- 보험계리사: 약간 불만족

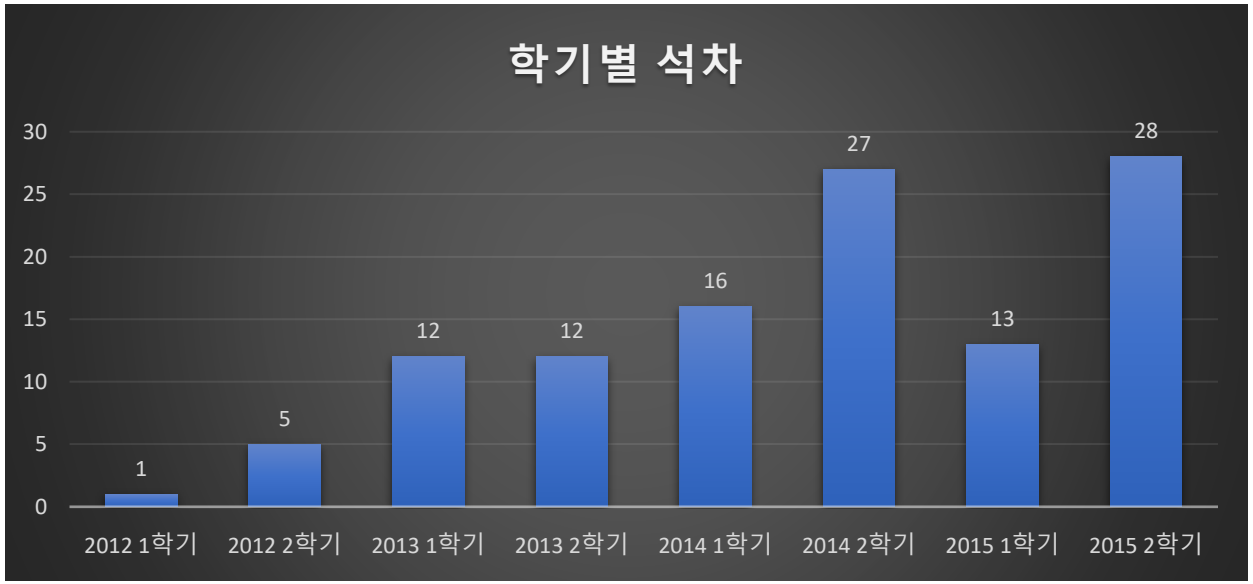
# Ch1. 정보통계 보험수리학과의 고종현



# 보다 구체적인 적성 찾기

- ✓ 인생에 대한 고민
- ✓ 나는 무엇이 될 것인가?

- 타과 (수학, 경영, 산업공학) 전공 수업만 48학점..
- 학점 관리의 어려움 vs 확실한 적성 발견




**인공지능 시스템 (산업공학) 수업 중 인공지능 시스템 연구실 (산업공학)에 스카우트된 사연:**

***<스팸 메시지 필터링 알고리즘 및 시스템 개발>***

- Case Based Reasoning 기반의 머신러닝 알고리즘을 통한 효과적인 스팸 메시지 EDA 분석 및 예측
- 웹 기반 시스템 개발을 통한 실제 사용환경 구현

**"본인이 잘하고 자신감 있는 분야를 찾는 것이 중요"**

# Ch2. 연구자로서의 길

## 현실: 요즘 컴공과 출신 아니면 머신러닝/알고리즘 분야 채용안되는 이유..

- **기업에서의 AI 서비스 아이템 개발은 데이터 엔지니어 능력보다는, 시스템 엔지니어로서의 능력을 우선 요구함**
- **2017~2020년 마구잡이식 채용으로 발생한 기업의 편견 (C*** 기업: 기존에 채용한 컴공과 외 타학과 출신 개발자들을 분리시켜 새로운 부서로 배정하고, 앞으로 컴공과 출신만 뽑음)**
- **통계학과 출신의 경우, [데이터 분석가] 채용 분야로 지원할 시 컴공과 출신보다 특혜 존재**
- **카카오 개발자 채용 분야 예시:** 플랫폼 개발자, 서버 개발자, **데이터 분석가**, 데이터 사이언티스트, 데이터 엔지니어, DB 엔지니어, 안드로이드 개발자, Ios 개발자, 클라우드 관리자, 소프트웨어 개발자

- ✓ 통계학과의 장점: 딥러닝/알고리즘 분야보다는 **데이터 분석 분야** 지원을 추천
- ✓ 딥러닝/알고리즘 분야 대기업 취업을 원할 경우: AI대학원 및 컴퓨터공학 대학원이 베스트
- ✓ 딥러닝/알고리즘 분야 학계 진로를 원할 경우: AI대학원 및 수학/통계학과 대학원이 더 큰 차별성 존재

# Ch2. 연구자로서의 길

## 대학원에 가면 무엇을 할까? ( 개인연구:기업과제 = 9:1 비율인 연구실 )

### # 개인연구

- ✓ 산업계 적용 연구가 아닌, (novel, 세계 최초) 방법론적인 수준에서의 연구
- ✓ 지도 교수님과 매주 1회 개인연구 미팅

### # 정부&기업과제 참여

- ✓ 2016 ~ 2017: 인공지능 후각 머신 개발 (악취 및 유독성 가스 분류)
- ✓ 2018 ~ 2020: 부산대병원 로그 데이터 분석을 통한 외래환자 프로세스 최적화 및 대기시간 감소
- ✓ etc.

### # 박사후연구원이 된 후

- ✓ 2022 ~ : 정부과제 책임연구원 (국내 최초 탄소중립 디지털 트윈을 GreenBPM 플랫폼 개발 시작)

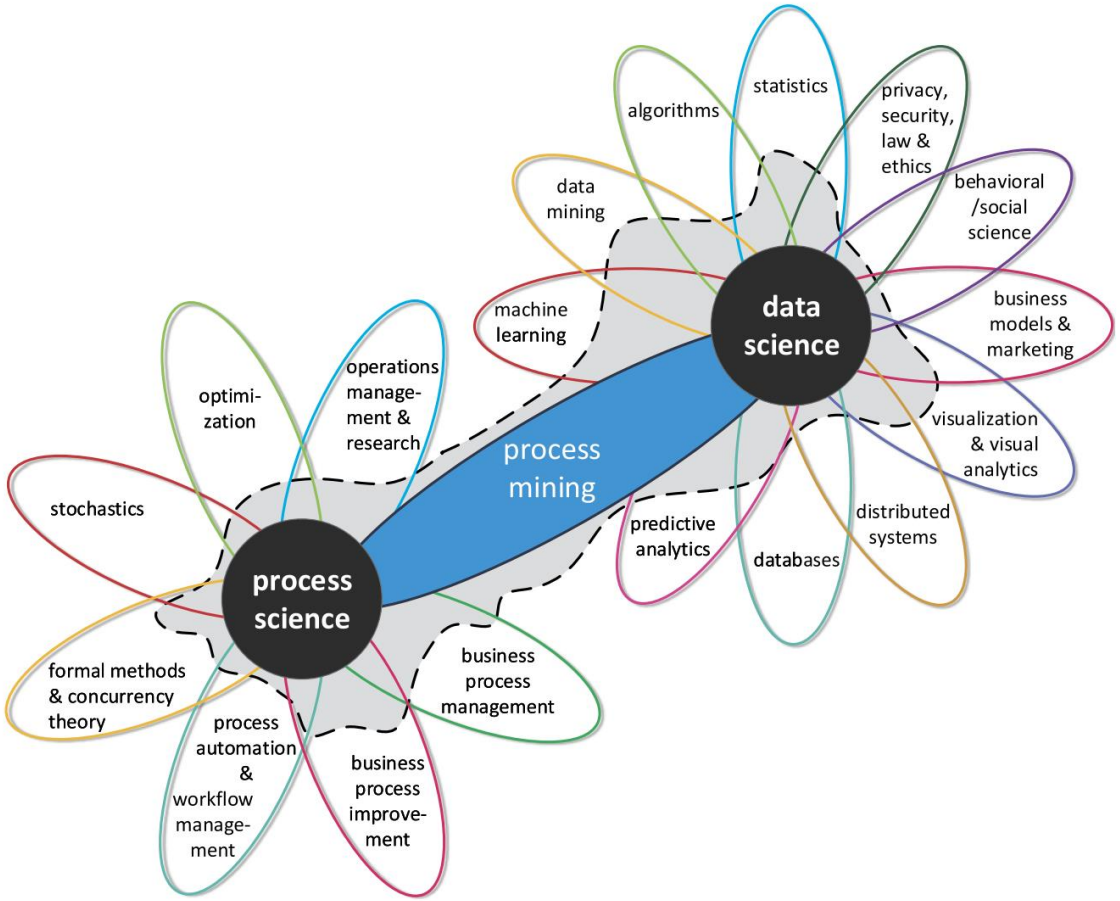
# Ch2. 연구자로서의 길

## AI 분야 취업 및 대학원에 관심이 있는 학생

- ✓ 2017년, 해외 학계는 이미 레드오션화 (반면, 국내는 이제서야 관심가지기 시작)
- ✓ 2022년, 국내 학계뿐만 아니라 취업 시장도 레드오션화 (카카오브레인의 경우 어림잡아 1000대1 이상의 채용경쟁률)

- ✓ 앞으로 이 분야에서 살아남기 위해선, 본인만의 특별한 무기가 있어야..

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝




*<프로세스 마이닝 창시자>*

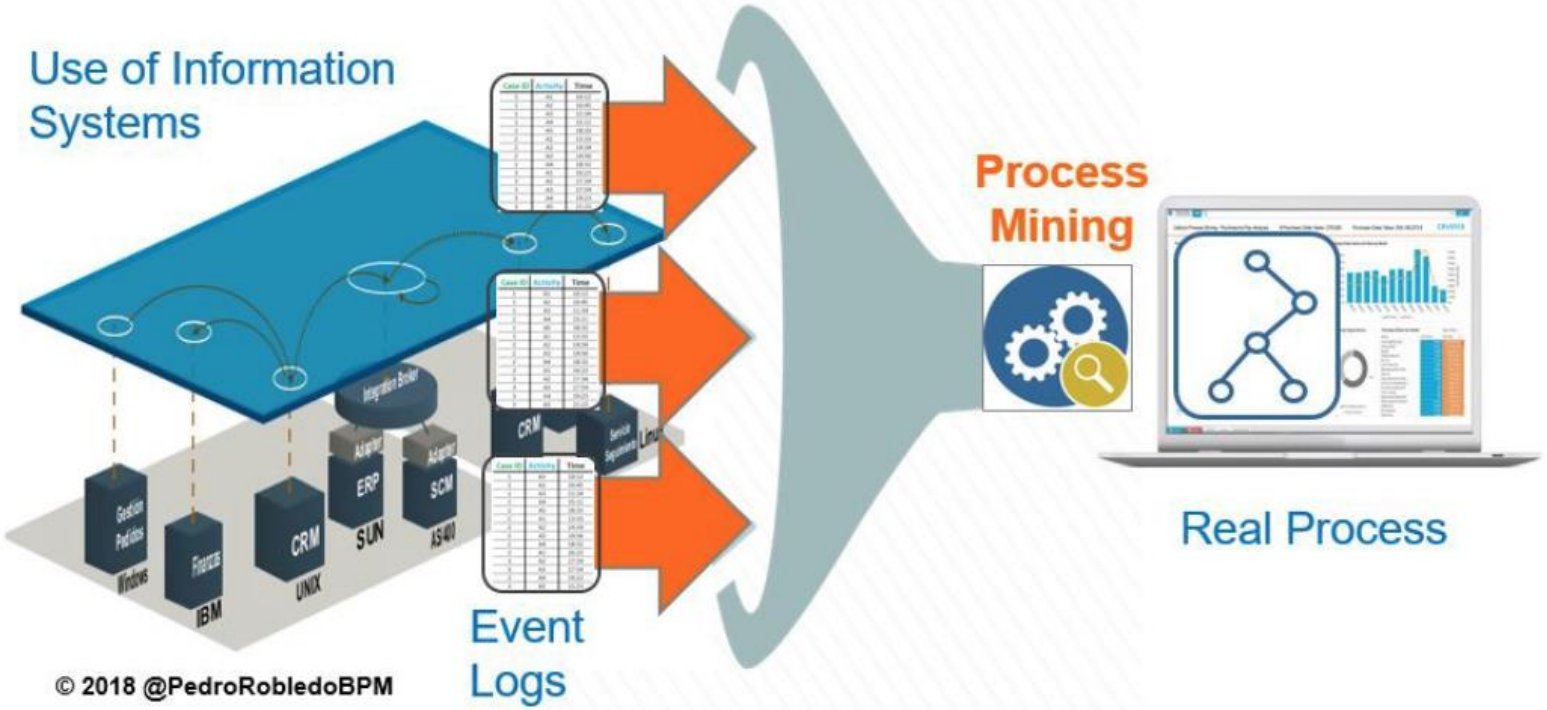
**Top Researchers by H-Index**
- **Computer Science and Electronics**

**(Ranking is based on Google Scholar H-Index. Released in March, 2018)**

**#14 : Wil M. P. van der Aalst**
**#53 : Yoshua Bengio**
**#88 : Yann LeCun**
**#291 : Geoffrey E. Hinton**

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

- ✓ 기관/기업의 모든 업무수행 관련 활동은 정보시스템에 의해 기록
- ✓ 이렇게 무한히 쌓이고 있는 기록 데이터를 '**이벤트 로그**' or '로그 데이터 ' 라 함



# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

✓ 전통적인 적용분야

- 업무 프로세스 개선, 병원 프로세스 개선 (Y증권, 분당서울대병원, 보라매병원, 삼성의료원)
- 전시회 동선 분석, 항만 물류 분석 (COEX, 부산항)
- 공정 프로세스 분석 (삼성전자, 삼성전기, LCD Display 제조 기업)
- 대출 프로세스 분석 및 사기 탐지 (해외 금융사)
- 온라인 예약 프로세스 분석 (부킹닷컴 네덜란드 본사 프로세스 마이닝 팀) Booking.com

## 무한한 적용성

- 스마트워치 등 각종 IoT 장비, 유튜브, 웹사이트 등 우리의 모든 일상에서 로그 데이터가 생성되고 있어, 앞으로의 프로세스 마이닝의 확장성은 무궁무진

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

- ✓ 기관/기업의 모든 업무관련 활동은 정보시스템에 의해 기록
- ✓ 이렇게 무한히 쌓이고 있는 기록 데이터를 '**이벤트 로그**' or '로그 데이터 ' 라 함

kakao enterprise

**[검색데이터] 대용량 로그 시스템 개발자 영입**

(주)카카오엔터프라이즈 | 경기

**주요업무**

- 로그레이크 플랫폼 개발
- 다양한 검색서비스에서 발생하는 로그 설계 및 수집 개발
- 대용량 데이터 정제 및 필터링 개발
- 대용량 데이터 실시간 모니터링 개발
- 대용량 데이터의 다양한 지표 개발

BNK경남은행 **(석사급 연봉 7천만원 이상)**

| | |
|---|---|
| **지원자격** | - 전공무관<br>- 데이터기반 분석 및 화면 설계 역량 보유<br>- 디자인 및 개발 협업 시 원활한 커뮤니케이션 역량<br>- 프로토타이핑 도구(Sketch, Adobe XD 등)를 활용한 산출물 제작 능력 보유<br>- 앱/웹 기반기술(그래픽, HTML&CSS 등)에 대한 기본 이해도 보유<br>- 사용자 리서치 및 로그 데이터 분석 등 고객 Needs 및 Pain Point 발굴 역량 보유<br>- IT개발/운영 전반 프로세스에 대한 경험 및 이해 |
| **주요업무** | - 디지털뱅킹 서비스 기획 및 운영<br>- 프로세스, 화면 세부 설계 및 관리<br>- 운영 서비스 모니터링 및 사후 효과 분석<br>- 모바일App 내 특화 서비스 제공전략 수립 및 운영 |

❖ **카카오 및 주요 국내 대기업**

**- 최근에서야 프로세스 마이닝 분야를 인지하였으며, 전문인력을 채용하고 싶지만 국내 전문인력이 턱없이 부족한 상태**

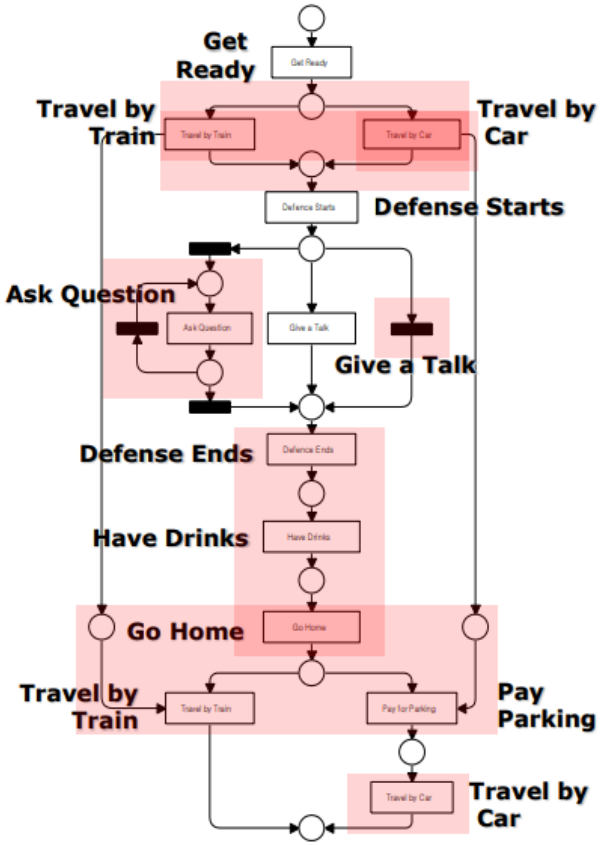
**- 현재 국내 프로세스 마이닝 전문 연구실은 단 3곳 (포항공대 송민석 교수, 유니스트 Marco Comuzzi 교수, 부산대학교 배혜림 교수)**

**- 카카오의 경우, 현재 시스템 로그 설계 단계로 프로세스 마이닝 연구 준비단계**

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

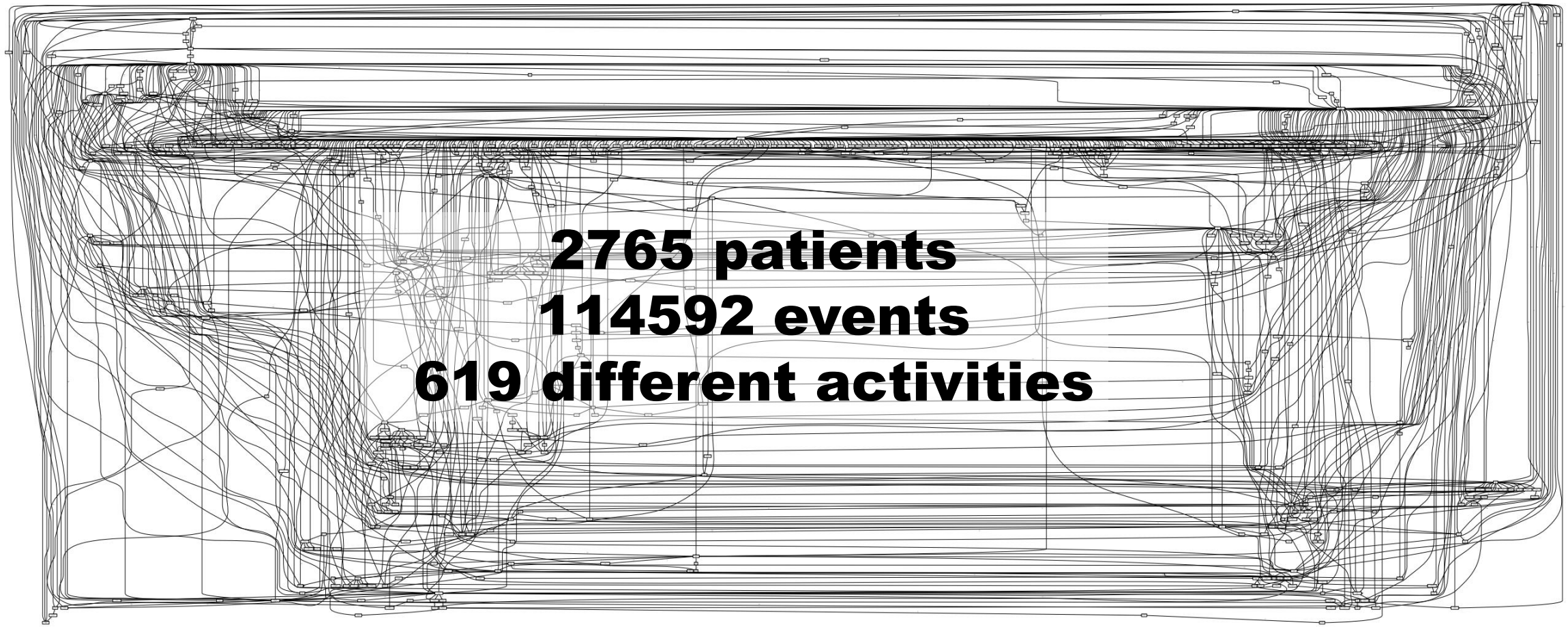
## ✓ 도전과제1: 이벤트 로그의 구조 이해

| case id | event id | properties | | | | |
|---|---|---|---|---|---|---|
| | | timestamp | activity | resource | cost | ... |
| 1 | 35654423 | 30-12-2010:11.02 | register request | Pete | 50 | ... |
| | 35654424 | 31-12-2010:10.06 | examine thoroughly | Sue | 400 | ... |
| | 35654425 | 05-01-2011:15.12 | check ticket | Mike | 100 | ... |
| | 35654426 | 06-01-2011:11.18 | decide | Sara | 200 | ... |
| | 35654427 | 07-01-2011:14.24 | reject request | Pete | 200 | ... |
| 2 | 35654483 | 30-12-2010:11.32 | register request | Mike | 50 | ... |
| | 35654485 | 30-12-2010:12.12 | check ticket | Mike | 100 | ... |
| | 35654487 | 30-12-2010:14.16 | examine casually | Pete | 400 | ... |
| | 35654488 | 05-01-2011:11.22 | decide | Sara | 200 | ... |
| | 35654489 | 08-01-2011:12.05 | pay compensation | Ellen | 200 | ... |
| 3 | 35654521 | 30-12-2010:14.32 | register request | Pete | 50 | ... |
| | 35654522 | 30-12-2010:15.06 | examine casually | Mike | 400 | ... |
| | 35654524 | 30-12-2010:16.34 | check ticket | Ellen | 100 | ... |
| | 35654525 | 06-01-2011:09.18 | decide | Sara | 200 | ... |
| | 35654526 | 06-01-2011:12.18 | reinitiate request | Sara | 200 | ... |
| | 35654527 | 06-01-2011:13.06 | examine thoroughly | Sean | 400 | ... |
| | 35654530 | 08-01-2011:11.43 | check ticket | Pete | 100 | ... |
| | 35654531 | 09-01-2011:09.55 | decide | Sara | 200 | ... |
| | 35654533 | 15-01-2011:10.45 | pay compensation | Ellen | 200 | ... |
| 4 | 35654641 | 06-01-2011:15.02 | register request | Pete | 50 | ... |
| | 35654643 | 07-01-2011:12.06 | check ticket | Mike | 100 | ... |
| | 35654644 | 08-01-2011:14.43 | examine thoroughly | Sean | 400 | ... |
| | 35654645 | 09-01-2011:12.02 | decide | Sara | 200 | ... |
| | 35654647 | 12-01-2011:15.44 | reject request | Ellen | 200 | ... |
| ... | ... | ... | ... | ... | ... | ... |



# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

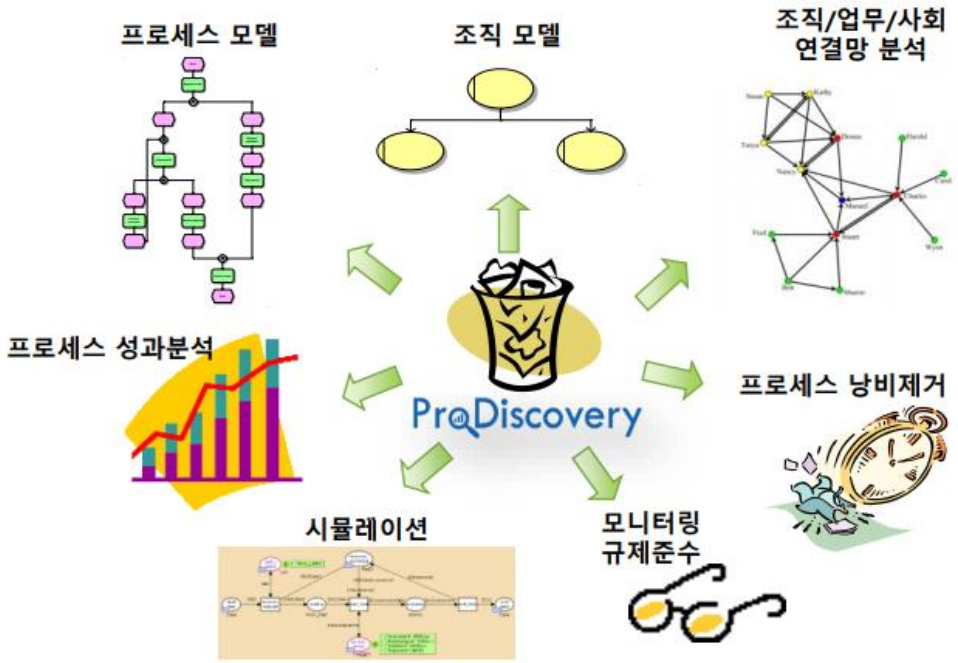
✓ 도전 과제2: 전체 프로세스 도출



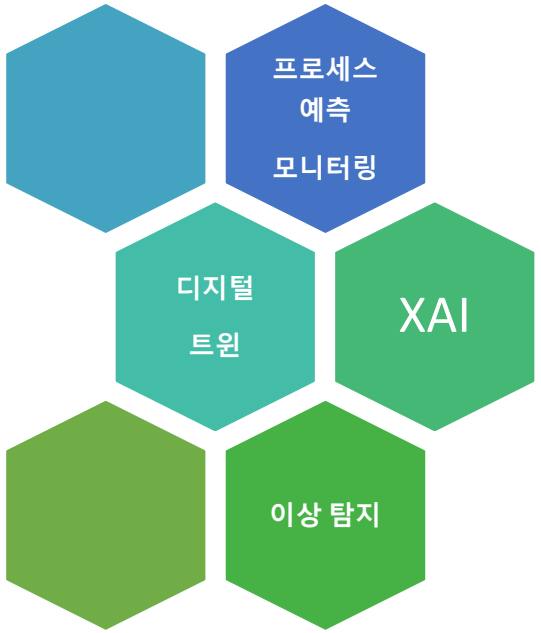
# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

✓ 도전과제3: 프로세스 분석

❖ 전통적인 프로세스 마이닝 기법 (~2016)




❖ 프로세스 마이닝 분야 최신 기법

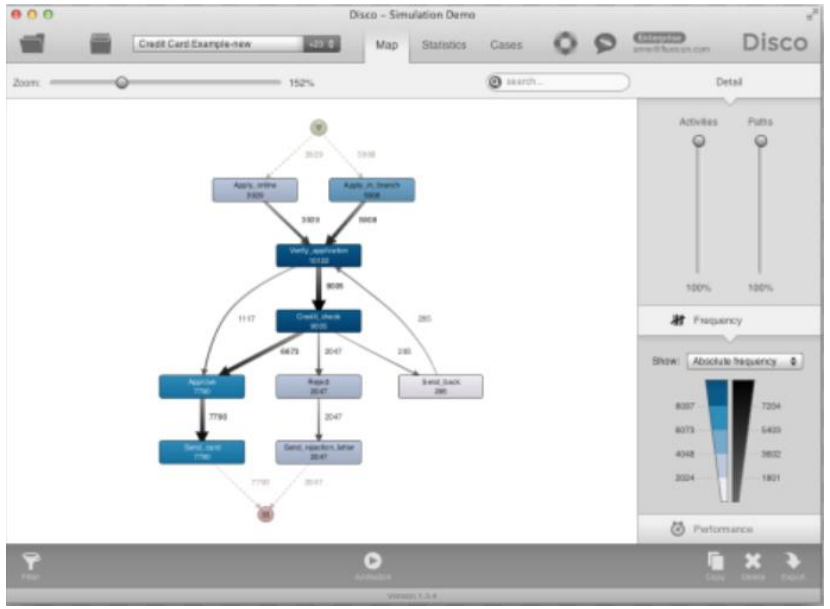


# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

✓ 프로세스 마이닝 패키지

ProM process mining workbench

Disco by Fluxicon




python

PM4PY

PROCESS MINING FOR PYTHON

bupaR

<통계 패키지>

SPSS

STATA

python

R

SAS

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

✓ 프로세스 마이닝 적용 사례1 : 프로세스 마이닝 기반 탄소 발자국 모니터링 기술

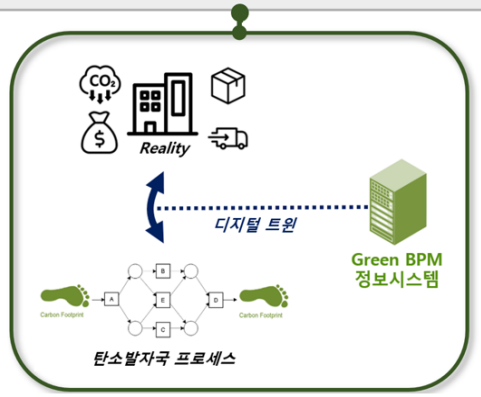
# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

## ✓ 프로세스 마이닝 적용 사례1 : 프로세스 마이닝 기반 탄소 발자국 모니터링 기술

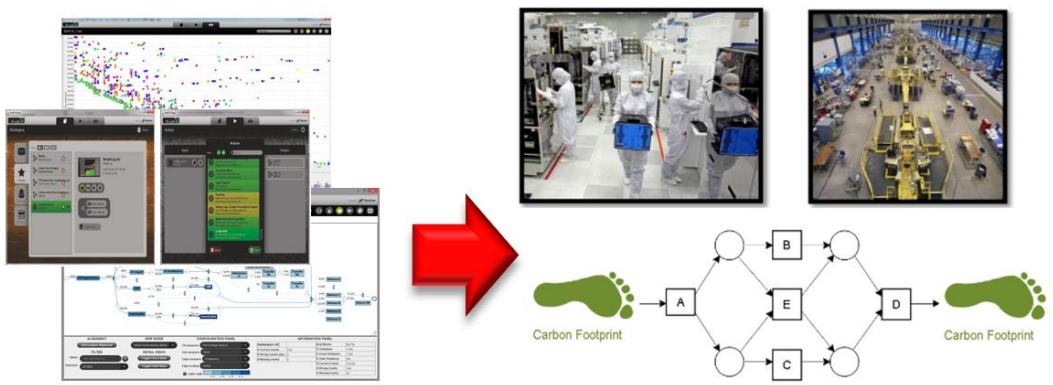
- 국내 최초 Green BPM (Business Process Management) 정보시스템 구축
- 프로세스 마이닝 기반 시뮬레이터를 통한 디지털 트윈 실현
  - 공정 프로세스 연동: 공정 프로세스 상 각 작업/장비 별 탄소 배출량을 산정하여 사업장·산업단지와 체계적 연계
  - 디지털 트윈 실현: 공정 프로세스의 최적화 및 변경에 따른 탄소 배출량 저감 모니터링
  - 데이터 거버넌스 활용: 로그 데이터를 활용한 AI 기반 배출원 이상상태 실시간 예측 모니터링 기술 개발

### GreenBPM 정보시스템

- 기업의 세부 프로세스 활동 별 탄소배출량 조사
- 탄소배출량 감소를 위한 프로세스 최적화 기술 적용 (친환경 기술 등을 포함한 모든 자원 운영 기술)




### 프로세스 마이닝 기반 탄소 발자국 프로세스 모델링



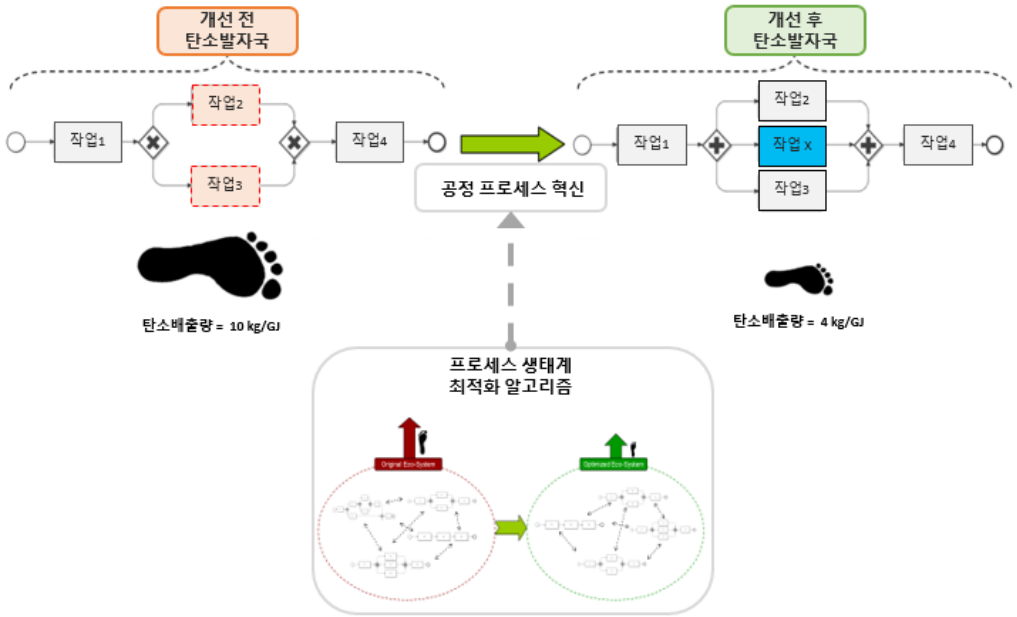

*프로세스 마이닝 기반 시뮬레이터 (ProM, PLG2, Disco 등)*

*탄소 발자국 프로세스 모델*

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

## ✓ 프로세스 마이닝 적용 사례1 : 프로세스 마이닝 기반 탄소 발자국 모니터링 기술

- 프로세스 마이닝 기반의 **탄소발자국 프로세스 모델** 구축 (*생산 건물 + 창고 건물로 구성된 **제조기업***)
- 탄소배출량 산정의 높은 정확도 (i.e., 94%)
- **탄소저감** 및 **공정 프로세스 혁신 (공정 가동률, 인력 최적화)** 동시 실현을 위한 **프로세스 최적화** 방법론 제시
- **전력/연료소비량 -14.8% 저감 실현**



# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

✓ 프로세스 마이닝 적용 사례1 : 프로세스 마이닝 기반 탄소 발자국 모니터링 기술

에너지 · 자원

## 탄소중립기본법 시행(2022. 3. 25.), 2050 탄소중립 법제화

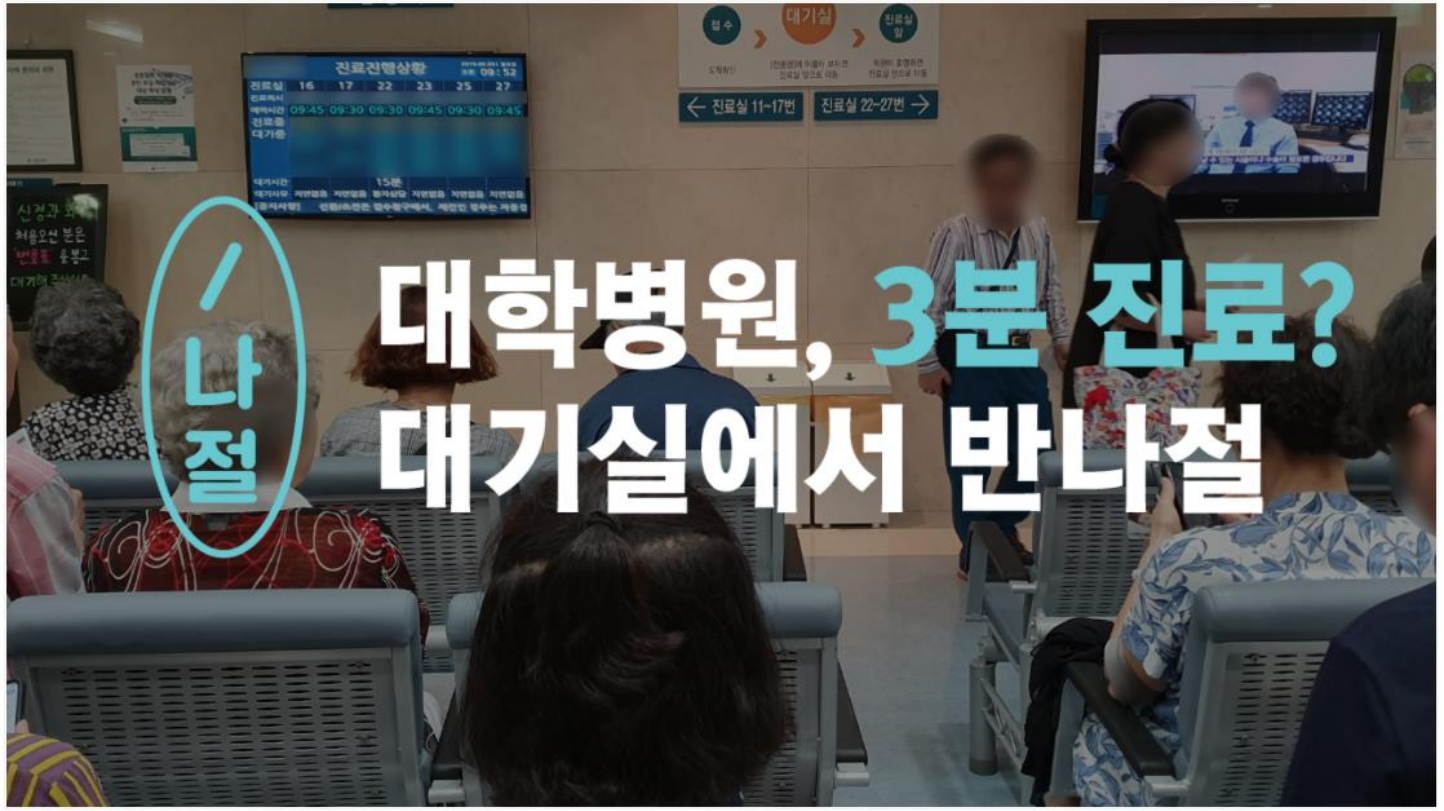
2022.04.04

▲ 기후변화영향평가 제도의 도입 대상 및 방법을 구체화
▲ 온실가스 관리업체의 지정 및 감축 목표 설정에 관한 사항을 구체화
▲ 2030년 중장기 국가 온실가스 감축 목표를 **2018년 온실가스 배출량 대비 40%**로 명시

대규모 규제가 시행될 것이 발표됨에 따라, 기업들이 앞다투어 프로세스 마이닝 기반의 GreenBPM 솔루션을 구축하려고 하는 상황

# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

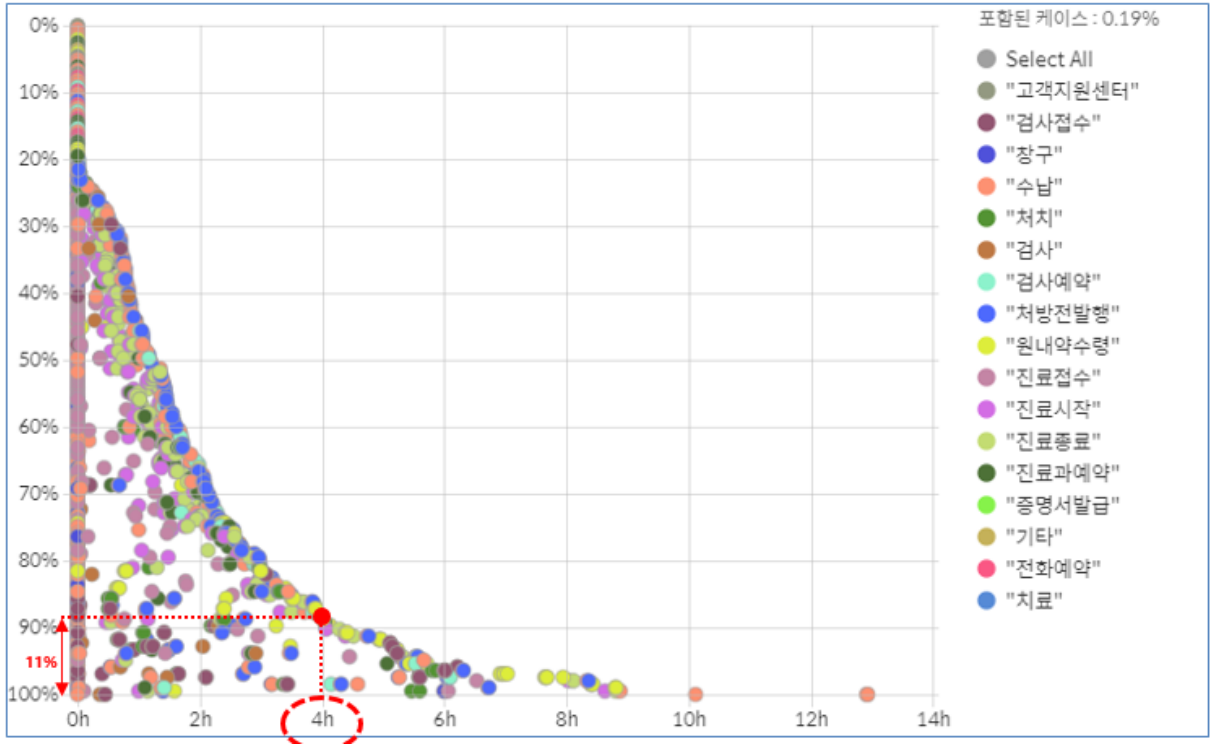
✓ 프로세스 마이닝 적용 사례2 : 병원 환자 대기시간 감소를 위한 진료 프로세스 개선



# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

## ✓ 프로세스 마이닝 적용 사례2 : 병원 환자 대기시간 감소를 위한 진료 프로세스 개선

- **소화기 내과 환자 대기시간 (가로축: 대기시간, 세로축: 환자수 누적 비율)**
  - 총 대기시간이 **4시간 이상인** 환자: 전체 환자 중 약 11%



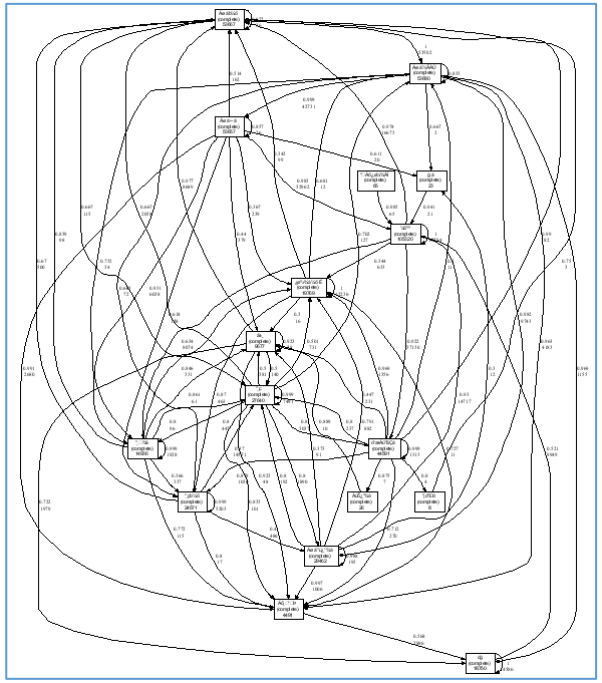
# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

## ✓ 프로세스 마이닝 적용 사례2 : 병원 환자 대기시간 감소를 위한 진료 프로세스 개선

<이벤트 로그 예시: 부산대학교병원 소화기내과>

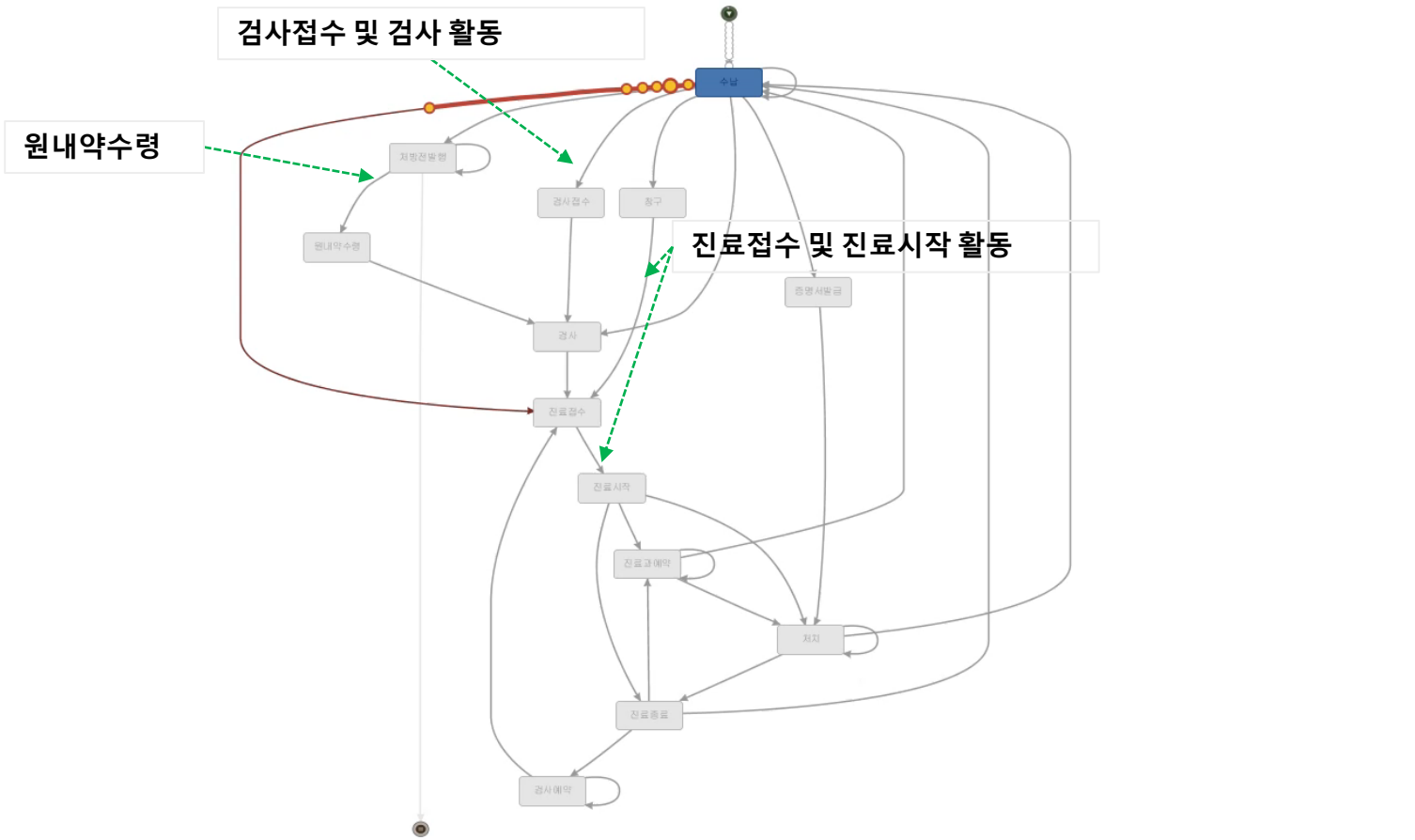
| 케이스ID | 활동 | 작업완료시간 | 작업자ID | 작업자부서 | 환자ID | 활동_version2 | 환자유형 |
|---|---|---|---|---|---|---|---|
| 1 | 진료과예약 | 201712201440 | 1930108 | 소화기내과 | 877 | 진료과예약/소화기내과 | 재진 |
| 1 | 수납 | 201712201443 | 2017909 | 원무팀 | 877 | 수납/원무팀 | 재진 |
| 1 | 처방전발행 | 201712201448 | KIOSK11 | 데이터없음 | 877 | 처방전발행/데이터없음 | 재진 |
| 2 | 진료과예약 | 201803141458 | 2008187 | 소화기내과 | 877 | 진료과예약/소화기내과 | 재진 |
| 2 | 수납 | 201803141516 | 2017909 | 원무팀 | 877 | 수납/원무팀 | 재진 |
| 2 | 처방전발행 | 201803141517 | KIOSK11 | 데이터없음 | 877 | 처방전발행/데이터없음 | 재진 |
| 3 | 진료과예약 | 201809271031 | 2008187 | 소화기내과 | 877 | 진료과예약/소화기내과 | 재진 |
| 3 | 수납 | 201809271046 | 2017908 | 원무팀 | 877 | 수납/원무팀 | 재진 |
| 3 | 처방전발행 | 201809271047 | KIOSK11 | 데이터없음 | 877 | 처방전발행/데이터없음 | 재진 |
| 3 | 검사예약 | 201809271104 | 5940401 | 원무팀 | 877 | 검사예약/원무팀 | 재진 |
| 3 | 검사예약 | 201809271105 | 5940401 | 원무팀 | 877 | 검사예약/원무팀 | 재진 |

<소화기내과 진료 프로세스 도출>



# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

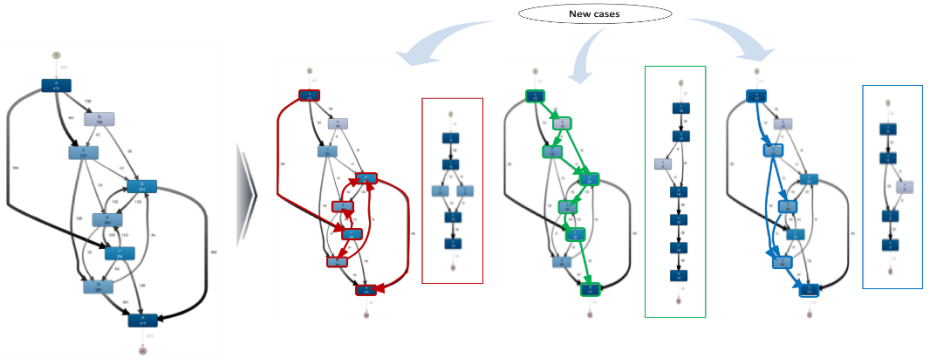
✓ 프로세스 마이닝 적용 사례2 활동 간의 병목 구간 탐색 (원내약발행, 진료접수, 검사 부분 병목 발생)



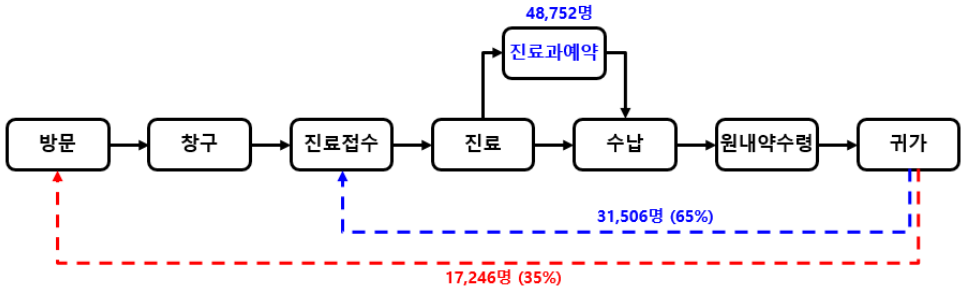
# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

## ✓ 프로세스 마이닝 적용 사례2 : 병원 환자 대기시간 감소를 위한 진료 프로세스 개선

- 프로세스 클러스터링 분석 (대기시간이 긴 집단 vs 짧은 집단)




- 결과: 안내 부족으로 불필요하게 원무팀을 재방문 하는 환자 집단이 존재 (평균적으로 대기시간 69분 증가)

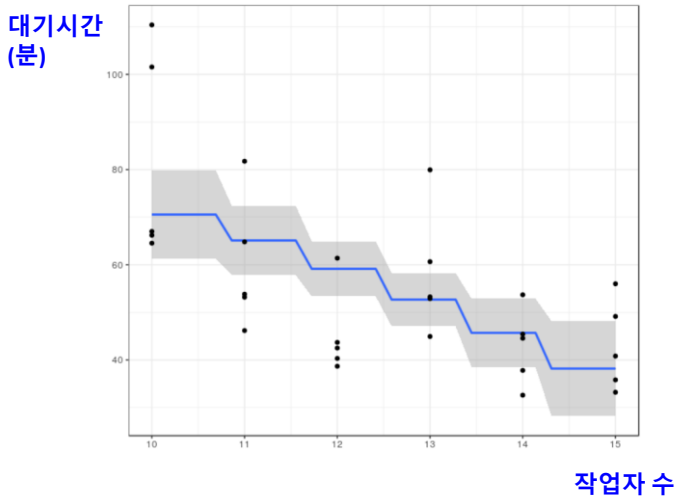


# Ch3. 새로운 데이터 사이언스 학문 분야, 프로세스 마이닝

## ✓ 프로세스 마이닝 적용 사례2 : 병원 환자 대기시간 감소를 위한 진료 프로세스 개선

- 작업자 수에 따른 대기시간 감소 **선형 모델 적합**
  → 작업자 수가 1명 증가할 때, 통계적으로 환자 1명 당 **7분** 정도의 대기시간이 감소
  → 소화기내과의 하루 평균 환자가 300명이라고 하였을 때, 1명 당 7분의 대기시간 감소면 매우 큰 경제적 효과

- **강화학습**을 적용한 예상 대기시간 예측 모델 개발
  → 실제 대기시간이 예측 대기시간보다 길 때, 환자의 불만 증가 : 학습 패널티 제공



# 마무리를 하며..

- ✓ 프로세스 마이닝 분야는 학계 및 취업 시장 모두 블루오션, 향후 최소 6년간 유지될 것
- ✓ 통계학과와의 시너지: 프로세스 마이닝은 궁극적으로 내부 로그 분석 및 프로세스 개선을 통해 기업의 수익 및 효율성을 극대화 하는 것으로, *데이터 분포에 대한 높은 이해도*와 *분석 스킬*을 가지고 있는 통계학과 출신들에게 적합
- ✓ 현재 세계 프로세스 마이닝 학계는 대부분 Computer Science 분야의 연구자들로 구성되어 있으며, 제 경우 통계학적인 지식을 차별화/강점으로 내세워 이미 학계에서 the Statistics Guy라는 별명이 붙음


- **Contact**

- Call: +82 10 7907 0013
- E-mail: whd1gus2@gmail.com
- Personal Page: http://sites.google.com/view/jonghyeonko